# Quantum Mechanical Study of Nanoscale MOSFET

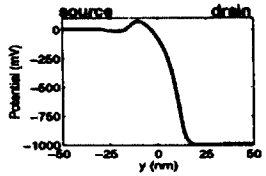
Alexei Svizhenko, M. P. Anantram, T. R. Govindan and Bryan Biegel

Nanoelectronics and Device Modeling Group
NASA Ames Research Center
Moffett Field, CA 94035-1000
U. S. A.



## Why do we need quantum simulations?

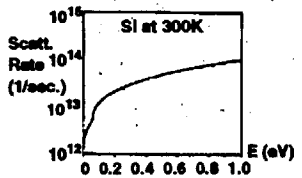
- Transit time through 25nm channel:
  $2.6\ 10^{-14}$ - $5.6\ 10^{-14}$ sec. (E=0.5eV)
  $5.8\ 10^{-14}$ - $1.3\ 10^{-13}$ sec. (E=0.1eV)
  (compare to scatt. rate on right column)




Experiments:

- Bell labs - 40nm; Transmittance ~ 0.8
- MIT-IBM-UCLA v-groove FET (10-35nm)

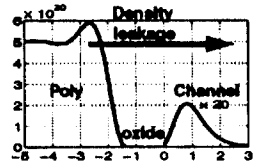
Lundstrom, Fund. of Carrier Transp.




- Location of inversion layer is important

  Quantitative modeling of $I_d$ versus $V_g$, $V_d$ is a challenge

- Gate leakage - Accurate modeling




- Benchmarking of classical (DD) and semiclassical (DG, BTE) simulators



## Approach

- Nonequilibrium Green's function equations and Poisson's equation are self-consistently solved in a non uniform grid (see figure).
- Details of equations and approximations necessary for the large scale simulations can be found in our preprint.
- Arbitrary doping profiles.
- Source, Drain and Gate are modeled as open boundaries.
- 2D gate leakage current is modeled.
- Anisotropic effective mass
- Discrete dopant effects are neglected.
- The results presented here are for the 'ballistic' case.

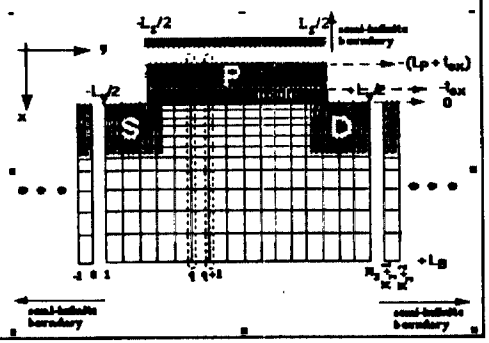





Fig. 1

## Results and Discussion

The steady state characteristics of MOSFETS that are of practical interest are the drive current, off-current, slope of drain current versus drain voltage, and threshold voltage. In this section, we show that quantum mechanical simulations yield significantly different results from drift-diffusion based methods. These differences arise because of the following quantum mechanical features:

(i) polysilicon gate depletion in a manner opposite to the classical case

(ii) dependence of the resonant levels in the channel on the gate voltage,

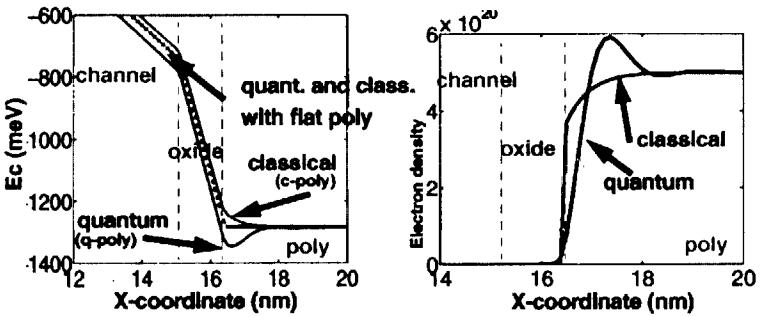
(iii) tunneling of charge across the gate oxide and from source to drain,

(iv) quasi-ballistic flow of electrons.

## Effect of polysilicon depletion region

The quantum mechanically calculated electron density near the $SiO_2$ barrier in the polysilicon region is smaller than the uniform background doping density. This is because the electron wavefunction is small close to the barrier. As a result, the potential profiles from drift diffusion calculations (c-poly) and the quantum calculations (q-poly) are different as shown in the figure below:

Fig. 2




Both profiles are identical in the subthreshold region, if polysilicon region is taken to have a flat band.



## Resonances and Transmission

In the absence of gate tunneling and inelastic tunneling, the quantum mechanical current is,

$$I_d = (2e/h) \int T_{SD}(E) [ f_S(E) - f_D(E) ]$$

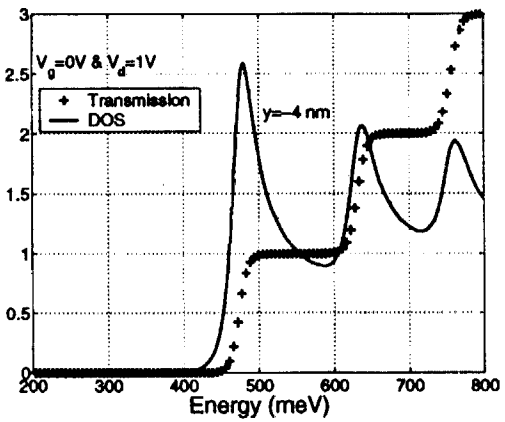
where, $T_{SD}(E)$ is the transmission probability from source to drain, and $f_S(E)$ and $f_D(E)$ are the Fermi-Dirac factors in the source and drain respectively. The total transmission (see figure 3) is step-like with integer values at the plateaus in-spite of the complicated two dimensional electrostatics. In visual terms, the energy at which the steps turn on are determined by an effective 'subband dependent' source injection barrier, in contrast to the source injection barrier in drift-diffusion calculations. This subband dependent source injection barrier is simply the maximum energy of the subband between source and drain due to quantization in the direction perpendicular to the gate plane (x-direction of Fig. 1). From a practical view point, the following two issues are important in ballistic MOSFETs:

(a) typically, the total transmission assumes integer value at an energy slightly above the maximum in 2D density of states

(b) the steps develop over 50 meV (twice the room temperature thermal energy). So, the shape of the steps is important in determining the value of current. Assuming a sharp step in total transmission with integer values in a calculation of current as in prior work is not quite accurate.



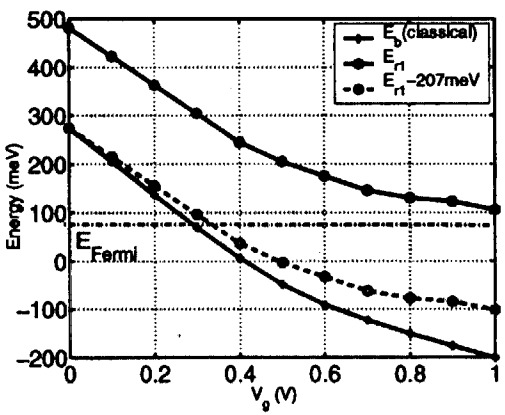
Fig. 3




We compare the results from our quantum simulations with published results from quantum-corrected Medici. To compare the quantum and classical results, an estimate of the energy of the first subband minima $E_{r1}$ from Fig. 3 and the location of the classical barrier height [$E_b$(classical)] are useful (see Fig. 4).



Fig. 4




Note that the slope is smaller in the case of $E_{r1}$ when compared to $E_b$(classical)

## Resonances and Transmission

In the absence of gate tunneling and inelastic tunneling, the quantum mechanical current is,

$$I_d = (2e/h) \int T_{SD}(E) [ f_S(E) - f_D(E) ]$$

where, $T_{SD}(E)$ is the transmission probability from source to drain, and $f_S(E)$ and $f_D(E)$ are the Fermi-Dirac factors in the source and drain respectively. The total transmission (see figure 3) is step-like with integer values at the plateaus in-spite of the complicated two dimensional electrostatics. In visual terms, the energy at which the steps turn on are determined by an effective 'subband dependent' source injection barrier, in contrast to the source injection barrier in drift-diffusion calculations. This subband dependent source injection barrier is simply the maximum energy of the subband between source and drain due to quantization in the direction perpendicular to the gate plane (x-direction of Fig. 1). From a practical view point, the following two issues are important in ballistic MOSFETs:

(a) typically, the total transmission assumes integer value at an energy slightly above the maximum in 2D density of states

(b) the steps develop over 50 meV (twice the room temperature thermal energy). So, the shape of the steps is important in determining the value of current. Assuming a sharp step in total transmission with integer values in a calculation of current as in prior work is not quite accurate.



Fig. 3

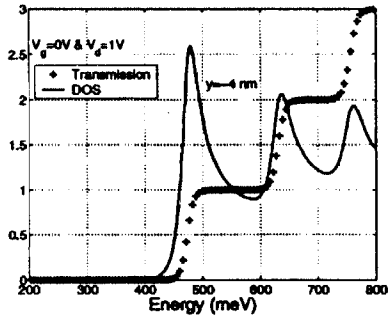
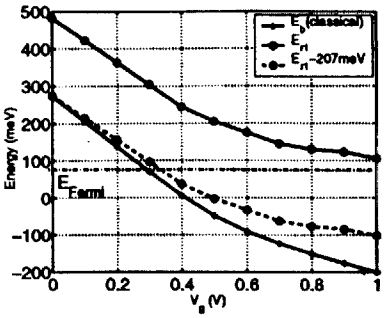



Fig. 4



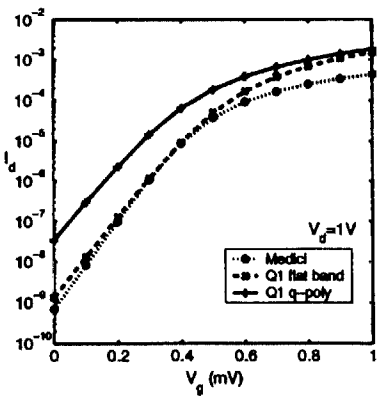

- Note that the slope is smaller in the case of $E_{r1}$ when compared to $E_b$(classical)

We compare the results from our quantum simulations with published results from quantum-corrected Medici. To compare the quantum and classical results, an estimate of the energy of the first subband minima $E_{r1}$ from Fig. 3 and the location of the classical barrier height [$E_b$(classical)] are useful (see Fig. 4).



## $I_d$ versus $V_g$

Fig. 5




Subthreshold region: The slope $d[\log(I_d)]/dV_g$ is smaller in the quantum case when compared to Medici (Fig. 5). Further, the current resulting from the simple intuitive expression

$$I = I_{q0} \exp(-E_{r1}/kT)$$

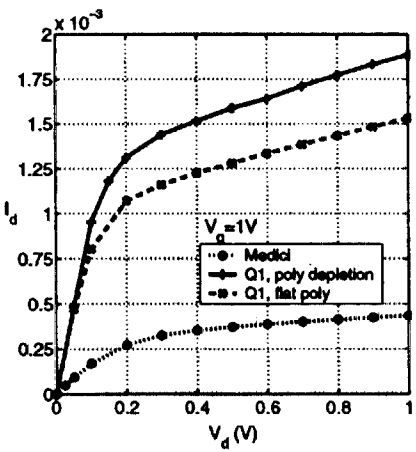
matches the quantum result quite accurately. $I_{q0}$ is a prefactor chosen to reproduce the current at $V_g$=0 in Fig. 5. The difference in slope between the classical and quantum results can be understood from the slower variation of $E_{r1}$ in comparison to $E_b$(classical) as a function of $V_g$ (Fig. 4) and the fact that $E_{r1}$ is well above the source Fermi energy

Large gate biases: The drain current and slope $d[\log(I_d)]/dV_g$ are larger in the quantum case. The higher $d[\log(I_d)/dV_g$ at large gate voltages in the quantum case can be understood from the fact that while $E_{r1}$ is above the Fermi level $E_b$(classical) is below, at $V_g$=1V [the quantum current is proportional to $\exp(-(E_{r1}-E_F)/kT)$]. The mobility model assumed in the classical case also plays a role in determining the slope.



## $I_d$ versus $V_d$

Fig. 6



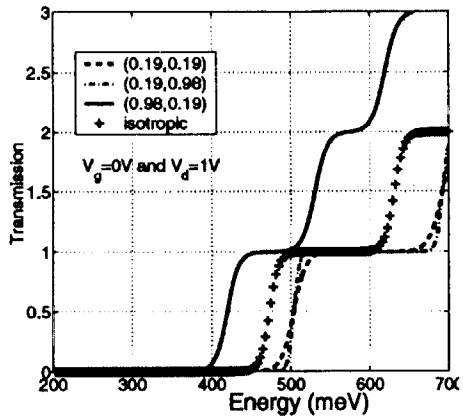

The values of $dI_d/dV_d$ and drive current are important in MOSFET applications because they determine switching speeds.

Figure 6 compares the drain current versus drain voltage for $V_g$=0 and $V_g$=1V. The drive current ($V_g$=1V) calculated using Q1 with the polysilicon region treated in the flat band and q-poly approximations is more than 100‰ and 200‰ larger than Medici results.

$dI_d/dV_d$ in the linear region is up to three times larger in Q1. The subthreshold drain current is smaller in Q1. We however expect that with decreasing channel length, the sub threshold $I_d$ will become larger than the Medici results due to quantum mechanical tunneling.

## Anisotropic Effective Mass

Fig. 7



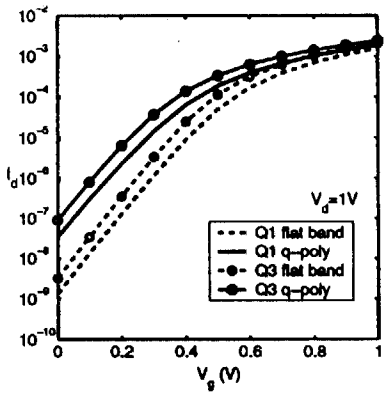

The primary influence of anisotropic effective mass is to change the energy of the subbands in the inversion layer.

Valleys with the largest effective mass perpendicular to the oxide (0.98 $m_o$) have subband energies that are smaller than the isotropic effective mass case.

We see from the plot of transmission versus energy (Fig. 7) that the valleys with ($m_x$ = 0.98 $m_o$, $m_y$ = $m_z$ = 0.19 $m_o$) have resonance levels that are more than 50 meV lower in energy than the isotropic effective mass case.



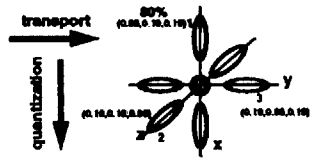
Fig. 8




The subthreshold current (Fig. 8) is a few hundred percent larger than the value obtained from the isotropic effective mass case. This follows by noting that the subthreshold current depends on exp(-$E_T$/kT). The drive current (Fig. 8) from the anisotropic effective mass case is more than twenty five percent larger than the isotropiceffective mass case.

The subthreshold current (Fig. 8) is a few hundred percent larger than the value obtained from the isotropic effective mass case. This follows by noting that the subthreshold current depends on exp(-$E_T$/kT).

Fig. 9



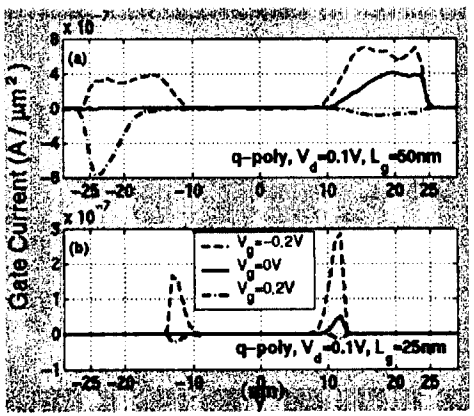

The drive current (Fig. 8) from the anisotropic effective mass case is more than twenty five percent larger than the isotropiceffective mass case.

We find that the valley with the largest $m_x$ (=0.98 $m_o$) carries 80% of the current at $V_g$ = $V_d$ = 1V. Thus all three valleys are necessary for an accurate calculation of the ballistic current.



## Gate Leakage Current

Fig. 10




A major problem in MOSFETs with ultra thin oxides is that tunneling from gate to drain will determine the off-current.

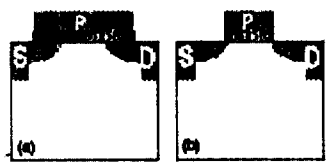
The gate leakage current versus y is plotted for the MIT25 device in Fig. 10. At $V_g$=0 V and $V_d$=1 V, the main path for leakage current is from the polysilicon gate contact on top of the oxide to the highly doped $n^+$ regions associated with the drain (Source Drain Extension, SDE) due to the large density of states in the highly doped region.

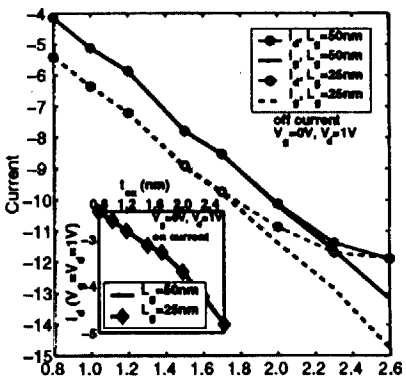
At non zero $V_g$, there is also an appreciable tunneling from the highly doped $n^+$ regions near the source to the polysilicon region on top of the gate.

We propose that gate leakage current can be significantly reduced without an accompanying reduction in drive current, by using shorter gate lengths. Any changes that result in a reduction of the gate leakage current should not alter the location of the resonant level at the source injection barrier (and hence the drive current).



Fig. 11



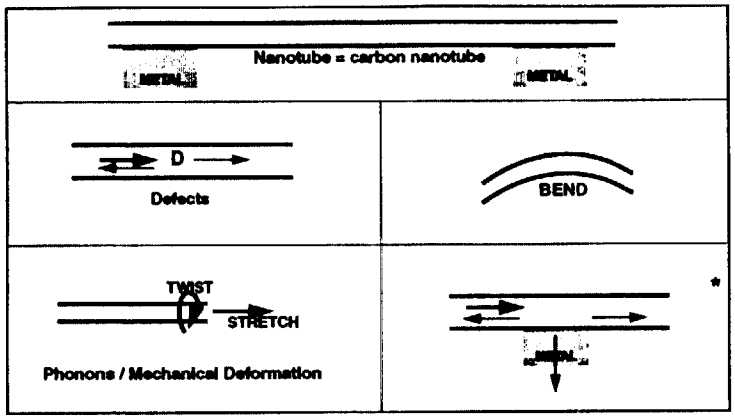

The main feature of the shorter gate lengths is a small overlap between the polysilicon gate and the n$^+$ region near the drain. This is pictorially represented in Figs. 11 (a) and (b) with 'long' and 'short' gate lengths.

To simulate the long and short gate lengths, we consider the doping profile of MIT25 with $L_g$ = 25 nm and 50 nm.

The off-current and gate leakage current are plotted in Fig. 12. We see that the gate leakage current reduces by more than an order of magnitude, and the drive current (see inset of Fig. 12) is within two percent of the $L_g$ = 50 nm case, as desired.

Though the gate leakage current reduces significantly, a drawback of this scheme is the requirement for very short polysilicon gate lengths. A polysilicon gate placed asymmetrically with respect to y=0 such that its overlap with the $n^+$ regions near the drain is small, will also serve to reduce the off-current without compromising the drive current.

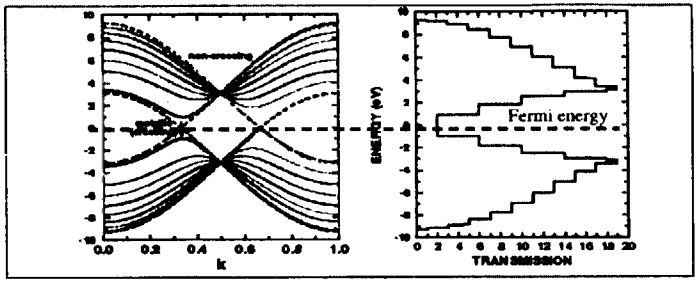
## Topics Studied




* Bragg reflec.: Intrinsic mechanism, which exists even in an ideal situation

* = This Poster

## Current-carrying capacity of carbon nanotubes

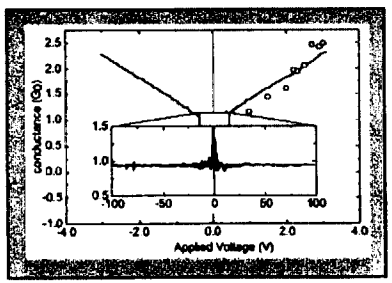



- Close to E=0, only two sub-bands, $Conductance = \frac{4e^2}{h}$ (6 kΩ)
- At higher energies, $Conductance = \frac{(20-30)e^2}{h}$ (< 1kΩ)

An important question is if subbands at higher energies be accessed to drive large currents through these molecular wires?

Experiments by various groups have shown that the differential resistance of a nanotube decreases by small amounts with increase in applied voltage. i.e., the current carrying capacity does not increase better than an ordinary resistor with applied bias.

One exception is the experiment by Frank et. al in Science 280 (1998) which showed a modest decrease in differential resistance with bias.

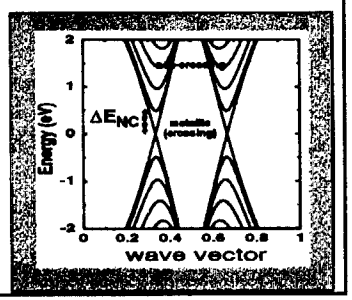



Main experimental features are:

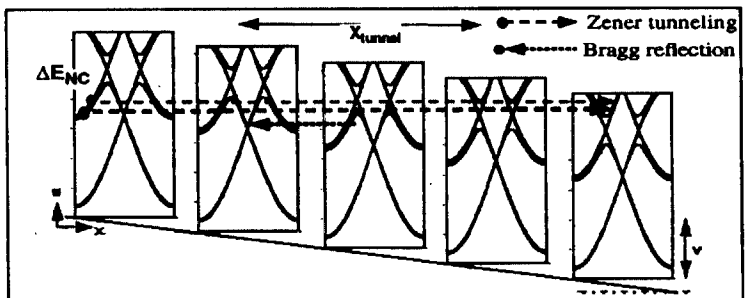
- $V_{APPLIED}$ < 200mV, conductance = $G_0$
- $V_{APPLIED}$ > 200mV, conductance increases modestly to about 1.75 $G_0$

- For large diameter nanotubes such as used in the experiments, the non-crossing subbands open at about 100 meV.
- Further there are about 80 subbands at an energy of about 2 eV.

- Yet the conductance is only ~ 1.75 $G_0$



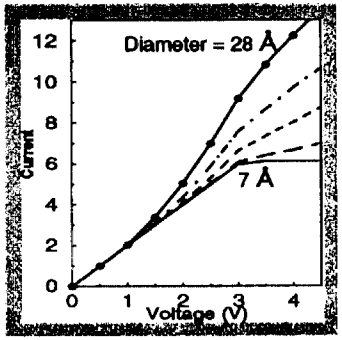

At large applied biases phonon scattering is usually important. Yet, the increase in differential resistance with bias is surprising. A semiclassical picture of electron flow from source (left) to drain (right) in the presence of a bias is useful:




An electron injected in the red crossing band is transmitted to the right. An electron injected in the green non-crossing band can either be Bragg reflected or can Zener tunnel to the right.

- The strength of the two processes are determined by:
  - Tunneling distance ($X_{tunnel}$)
  - Barrier height, $2\Delta E_{NC}$
  - Scattering and Defects

## Calculation of I-V characteristics in the ballistic limit




The main features are:

- $dI/dV \sim 4e^2/h$ for $V_a < 2\Delta E_{NC}$
- $dI/dV$ increases for $V_a > 2\Delta E_{NC}$
- Barrier height ($\Delta E_{NC}$) decreases with increase in diameter
- Total Current increases with increase in diameter

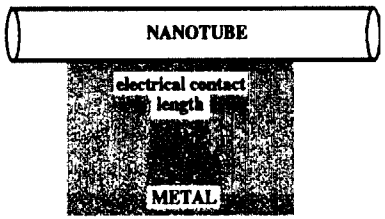
- The differential conductance is NOT comparable to the increase in the number of subbands. (For a (20,20) nanotube, there are 35 subbands at E = ± 3.5V.)
- Bragg reflection severely limits the current carrying capacity
  - crossing subbands are conducting
  - non-crossing subbands are primarily non-conducting except for a Zener tunneling component whose importance increases with increase in daimeter

## Coupling of carbon nanotubes to metallic contacts

The electronic properties of nanotubes are closely related to chirality. The metallic versus semicondting nature of nanotubes and the bandgap change with deformation depend on chirality.

Questions:

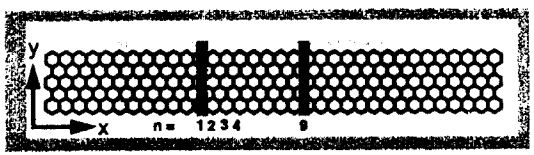
- Is there a preferable nanotube chirality to maximize current flow?
- Role of wave vector conservation?
- Explain experimentally observed scaling of conductance with contact length




Parameters that influence current flow:

- Strength of coupling to metal
- Length of metal-nanotube contact
- Defects
- Metal Fermi wave vector

## Scattering rate




The wave function of a nanotube is $\Psi = e^{ink_x L}\phi$

n is an integer and $\phi$ is wave func. of atoms in a 1D unit cell

The scattering rate from metal to nanotube (Born approx.) is,

$$1/\tau \; \alpha \; | < \Psi_{nt} \; | \; V_{m\text{-}nt} \; | \; \Phi_m > |^2$$

$$\delta(k_x - k_x^m) \qquad | < \phi \; | \; V_{m\text{-}nt} \; | \; \phi_m > |^2$$

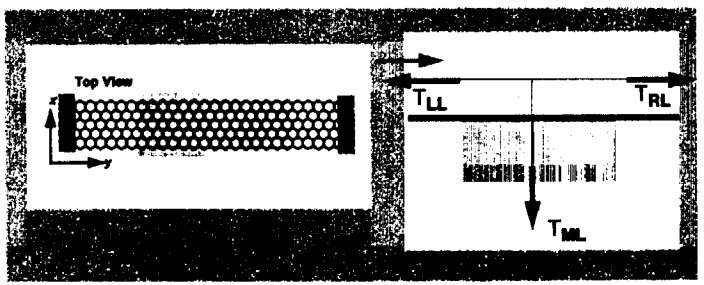
This implies that

- $k_x$ is conserved
- $k_y$ conservation is relaxed due to finite *width* of *contact area*

As a result of the difference in $k_y$ corresponding to E=0, important differences should arise as a function of chirality.

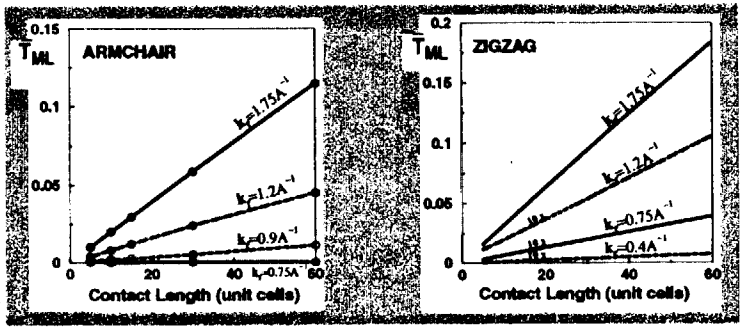
## How do we model the system?

- $\pi$ electron tight binding model
- Metal is modeled as a free electron gas ($k_F$)




- $T_{RL} + T_{ML} + T_{LL} = 2$.
- Phys. Rev. B, v.58, p. 4882 (1998) and v. 61, p. 14219 (2000)
- Compute self energy due to: (i) metal & (ii) semi-infinite CNT leads

Calculation of the total transmission from nanotube to metal, $T_{ML}$ are presented for armchair and zigzag nanotubes:



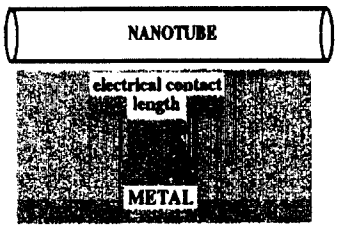

- Note that threshold value is the value of $k_f$ below which the transmission does not scale with contact length

The main differences between the armchair and zigzag case are:

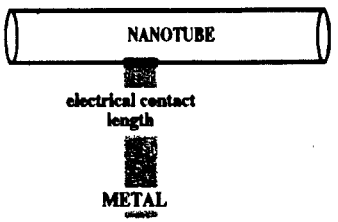
- Threshold value of $k_f$ is $\frac{2\pi}{3a_0}$ =0.85 Å$^{-1}$ for armchair nanotubes (see $k_f$ = 0.75Å$^{-1}$ ) and is 0 for zigzag nanotubes (see $k_f$ = 0.4Å$^{-1}$ ).
- Beyond the threshold $k_f$, transmission *increases* with contact length as seen in experiment by Tans et. al., Nature, vol. 386, 474 (1997)
- For zigzag tubes, $T_{LM}$ is small for $k_f \leq 1.2$ Å$^{-1}$ as a result of the large angular momentum in the circumferential direction.

The previous values of T are small compared to the maximum possible value of 2. Two possible scenarios to increase T are:

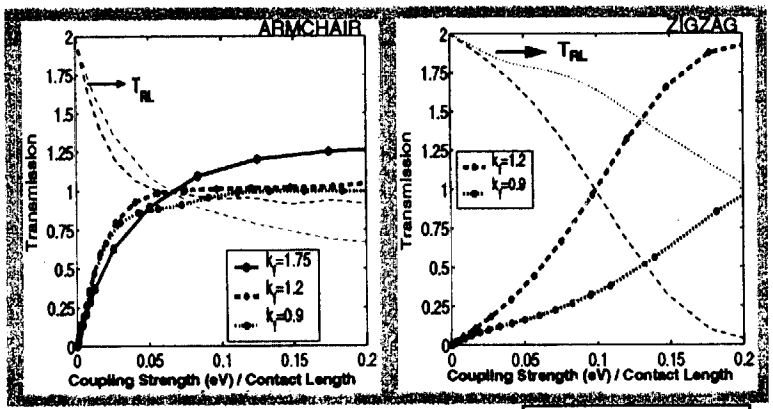
* Large contact length - Small coupling



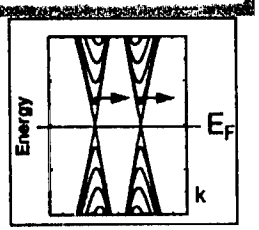

* Small contact length - Large coupling




- For nanoelectronics, the second option (right) is better.
- We model a contact length of *30 unit cells* (72 Å for armchair and 125 Å for zigzag nanotubes), and vary the coupling strength. The main results of the calculation are rather surprising and are presented below:



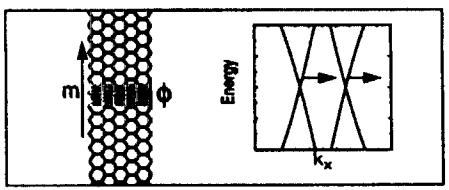

- Armchair:Transmission is pinned at values close to unity for metal $k_f$ of 0.9 and 1.2 Å$^{-1}$
- Zigzag: Transmission increases to two (maximum possible value)
- At small coupling strengths, transmission increases more rapidly in the armchair case

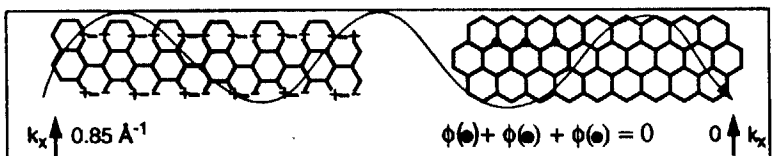



Metallic-zigzag are preferable from metal-nanotube coupling viewpoint

Nodes on the cylinder - Shape of NT wave function

$$\Psi = e^{imk_x a_0 \phi}$$




metal $k_f$=0.90 Å$^{-1}$ → distance between nodes 6 Å
Gold $k_f$=1.20 Å$^{-1}$ → distance between nodes 3.4 Å
Aluminum $k_f$=1.75 Å$^{-1}$ → distance between nodes 2 Å

Scattering rate from Metal to Nanotube $\alpha$ $\langle \phi_{nanotube} | H_{coupling} | \psi_{metal} \rangle$ $\langle + - | H_{coupling} | \psi_{metal} \rangle \sim 0$. Hence, Transmission is pinned close to unity.

- Side-contacted: zigzag nanotube are more desirable (curvature)
- Larger metal Fermi wave vector helps.

## Conclusions

- dI/dV versus V does not increase in a manner commensurate with the increase in number of subbands.


- The increase in dI/dV with bias is much smaller than the increase in the number of subbands - *a consequence of bragg reflection*
- Requirement for axial wave vector conservation:

| ARMCHAIR | ZIGZAG |
|---|---|
| cut-off $K_{Fermi}=2\pi/3a_0 \approx 0.85 Å^{-1}$ | cut-off $K_{Fermi} = 0$ |

- Our calculations show an increase in transmission with length of contact, as seen in experiments.
- It is desirable for molecular electronics applications to have a small contact area, yet large coupling. In this case, the circumferential dependence of the nanotube wave function dictates:
  - Transmission in armchair tubes saturates around unity
  - Transmission in zigzag tubes saturates at two